A Projection–Regularized Newton method for Nonlinear Ill–Posed Problems and its Application to Parameter Identification Problems with Finite Element Discretization [1]

Barbara Kaltenbacher

Abstract. This paper is concerned with nonlinear ill-posed operator equations $F(a) = y$ (e.g., parameter identification problems) and their approximate solution by a Newton type method that is regularized by projecting the linearized equation in each Newton step onto a finite-dimensional space (e.g., a finite element space), and by stopping the Newton iteration at an appropriate index. We prove convergence as the iteration index n goes to infinity in the noise free case and convergence as the data noise level δ goes to zero in the case of noisy data, as well as convergence rates under certain additional conditions.

Key words. Nonlinear ill–posed problems, iterative reguarization methods, stopping rule, convergence rates, parameter identification

AMS(MOS) Subject Classifications: 35R30, 65J15, 65J20, 47H17

Short title: Projection–Regularized Newton method

1 Introduction

The identification of a distributed parameter in a partial differential equation turns up in many applications in science and industry and is a typical example of a nonlinear ill–posed problem. Consider, e.g., the problem of identifying a distributed parameter $a = a(x)$ in the PDE

$$\begin{aligned} -\nabla(a\nabla u) &= 0 \qquad \text{in } \Omega \\ u &= g \qquad \text{on } \partial\Omega \end{aligned} \tag{1}$$

from measurements u^δ of u. As a model for such a nonlinear inverse problem we consider the abstract operator equation

$$F(a) = y \tag{2}$$

where $F : \mathcal{D}(F)(\subseteq X) \to Y$ is a nonlinear differentiable operator between two Hilbert spaces X and Y. (In the parameter identification problem (1) F is the parameter-to-solution map.) We are interested in the situation that the problem under consideration is ill–posed in the sense that a solution $a^\dagger$ does not depend continuously on the right–hand side y. Since in practice only approximate data with some error level δ, i.e.

$$\|y^\delta - y\| \le \delta \tag{3}$$

are available, problem (2) has to be regularized (see e.g. [8], [12],[22], [23], [26], [27], [28]). Throughout this paper we assume that a solution $a^\dagger$ of (2) exists, i.e.

$$F(a^\dagger) = y. \tag{4}$$

It is characteristic for ill–posed problems that the approximations to $a^\dagger$ produced by any regularization method converge to $a^\dagger$ — if at all — only arbitrarily slowly and that assertions on convergence rates are only possible under additional assumptions, so–called source conditions

$$a_0 - a^\dagger = \left(F'(a^\dagger)^* F'(a^\dagger)\right)^\nu w \quad \text{for some } w \in \mathcal{N}(F'(a^\dagger))^\perp \tag{5}$$

[1]Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under grant 10866 and the SFB F013/08

on the difference between the initial guess a_0 and the solution $a^\dagger$. Since the linear operator $F'(a^\dagger)$ is typically smoothing, they can often be interpreted as smoothness or regularity assumptions on the initial error $a_0 - a^\dagger$. In the nonlinear situation the norm of w in the source condition has to be sufficiently small, which reflects the fact that the analysis of solution methods also for well–posed nonlinear problems are usually based on a local analysis working with some closeness condition such as

$$\|a_0 - a^\dagger\| \leq \rho \tag{6}$$

with ρ sufficiently small, and globalization strategies have to be developed additionally (which is a task we do not consider here). Usually one has to take into account not only data error (3) but also the fact that the forward operator F cannot be evaluated exactly but has to be numerically approximated, i.e., one has, instead of F some F^ϵ. For parameter identification problems the forward operator F corresponds to the solution of a PDE, F^ϵ to its numerical approximation, e.g., by the finite element method.

Certainly the most well–known and most widely used regularization method for nonlinear ill–posed problems is Tikhonov regularization (see e.g. [8] and the literature cited there). In order to avoid the problem of finding a global minimizer of the Tikhonov functional, iterative regularization methods for nonlinear ill–posed problems have been developed, e.g. a nonlinear version of Landweber iteration (cf. [17]). Due to its superior convergence properties for well–posed problems Newton's method has been applied also to several inverse problems and recently some investigations on possible adaptions of Newton's method to ill–posed problems and on their convergence have been carried out (see e.g. [1], [3], [2], [19], [20], [15], [16], [24]).

In the analysis of all of these regularization methods for nonlinear ill–posed problems certain conditions on the degree of nonlinearity of F play an important role. It turns out that the usual Lipschitz condition on the Frechet derivative

$$\|F'(\bar{a}) - F'(a)\| \leq L\|\bar{a} - a\| \qquad \bar{a}, a \in \mathcal{B}_{2\rho}(a_0) \tag{7}$$

is too weak as long as no source condition or only (5) with $\nu < \frac{1}{2}$ holds (cf. Section 11.1 in [8]; see also [5]). In that situation one works with conditions that are similar or at least related to

$$F'(\bar{a}) = R_a^{\bar{a}} F'(a) \qquad \bar{a}, a \in \mathcal{B}_{2\rho}(a_0) \tag{8}$$

for some bounded linear operators $R_a^{\bar{a}}$ with $\|R_a^{\bar{a}} - I\| \leq C_R\|\bar{a} - a\|$, or

$$\|F(\bar{a}) - F(a) - F'(a)(\bar{a} - a)\| \leq \eta\|F(\bar{a}) - F(a)\| \qquad \bar{a}, a \in \mathcal{B}_{2\rho}(a_0)\ , \tag{9}$$

(which by the way locally follows from (8)) and that also have been verified for some applications, e.g. certain parameter identification problems (cf. [17]). In case (8), or more generally (9) together with

$$A_0 = RF'(a^\dagger) \tag{10}$$

for some bounded linear operators A_0, R with

$$\|I - R\| < 1 \tag{11}$$

holds, one can use a frozen Newton method in the methods proposed in [3], [2], [19], [20], and also in this paper. Note that neither of the operators on the right hand side of (10) needs to be known explicitely and that for some examples of parameter identification problems (10) can be established (cf. [17] and Section 3 below).

For a practical implementation, any regularization method has to be discretized. On the other hand, discretization, i.e., projection on a finite-dimensional space itself can have a regularizing effect: Finite dimensional problems — though they might be ill-conditioned — are always well-posed in the sense of stable dependence of the solution on the data, as soon as existence and uniqueness of a solution can be guaranteed (which at least in the linear case can always be done by an appropriate definition of a solution — e.g., the concept of a best-approximate solution). We use this fact to construct a Newton type method for (2) based on regularization by projection. In the following section we show its convergence as $n \to \infty$ in the noiseless case $\delta = 0$. In the situation of noisy data and/or inexact evaluation of F, due to the instability of the problem, the iteration has to be stopped after an appropriate number of steps in order to avoid an explosion of the propagated noise. We give an a priori and an a posteriori parameter choice strategy that yield convergence and, under a source condition, convergence rates as $\delta \to 0$. In the context of parameter identification problems for PDEs it is near at hand to use finite element for a discretization. We will show that this use of finite elements fits well into the framework of our projection–regularized Newton type method. Section 3 contains a simple model example of a parameter identification problem, for which the convergence statements are concretized. In Section 4 we verify the theoretical results of the foregoing sections in some simple numerical tests.

In the following we will use the notation c or C, for positive constants that are typically "small" or "large" (in the sense of "not necessarily small") respectively; they can have different values whenever they appear in the text, but they are introduced in such a way that no confusion should occur.

2 A projection–regularized Newton method. Convergence analysis.

As a convergence analysis in the linear case — see e.g. Section 3.3 in [8] — has shown, there are good reasons for using regularization by projection in image space rather than in preimage space. For an ill-posed linear equation

$$Bs = b \tag{12}$$

with $B : X \to Y$, X, Y Hilbert spaces, this method of regularization by projection (see [11], [7], [10], [13] — or, as Natterer ([25]) calls it, the dual least squares method,) is defined as follows: Let $\{\psi_1, \psi_2, \psi_3, \ldots\} \subseteq \mathcal{R}(B)$ be linearly independent with linear span dense in $\overline{\mathcal{R}(B)}$. As k-th approximation s_k for s we use the (unique) solution of the (k-dimensional) problem

$$\begin{array}{ll} \langle Bs, \psi_j \rangle = \langle b, \psi_j \rangle & j \in \{1, \ldots, k\} \\ s \in \operatorname{span}\{B^*\psi_1, \ldots, B^*\psi_k\} & \end{array} \tag{13}$$

or equivalently

$$s_k = (r_k B)^\dagger r_k b\,,$$

where † denotes the generalized inverse and

$$\begin{array}{rccc} r_k & : & Y & \to & \mathbb{R}^k \\ & & y & \mapsto & \begin{pmatrix} \langle y, \psi_1 \rangle \\ \vdots \\ \langle y, \psi_k \rangle \end{pmatrix} \end{array}$$

Applying this regularization method to the linearized equation

$$A_n(a_{n+1} - a_n) + F(a_n) - y^\delta = 0$$

in each Newton step we get the following Newton type regularization method for (2): Let

$$a_0 - a^\dagger \in \mathcal{N}(F'(a^\dagger))^\perp \ ,$$

and (6) hold and let, for each n, A_n be an approximation for $F'(a^\dagger)$:

$$A_n := F'(a_n) \ , \tag{14}$$

if (7) holds, and

$$A_n :\equiv A_0 \ , \tag{15}$$

if (10) holds. Selecting in each Newton step a linearly independent normed family $\{\psi_1^n, \psi_2^n, \psi_3^n \ldots\}$ and a dimension $k = k(n)$ with

$$\operatorname{span}\{\psi_1^n, \ldots, \psi_{k(n)}^n\} \subseteq \operatorname{span}\{\psi_1^{n+1}, \ldots, \psi_{k(n+1)}^{n+1}\} \tag{16}$$

$$\forall v \in \operatorname{span}\{\psi_1^n, \ldots, \psi_{k(n)}^n\} \ : \ \sqrt{\sum_{j=1}^{k(n)} \langle v, \psi_j^n \rangle^2} \geq c_\psi \|v\| \ , \tag{17}$$

$$\forall v \in Y \ : \ \sqrt{\sum_{j=1}^{k(n)} \langle v, \psi_j^n \rangle^2} \leq C_\psi \|v\| \ , \tag{18}$$

with some constants c_ψ, C_ψ, independent of n, as well as

$$\overline{\mathcal{R}(F'(a^\dagger))} \subseteq \overline{\bigcup_{n \in \mathbb{N}} \operatorname{span}\{\psi_1^n, \psi_2^n, \psi_3^n, \ldots\}} \, , \tag{19}$$

$$\{\psi_1^n, \ldots, \psi_{k(n)}^n\} \subseteq \overline{\mathcal{R}(A_n)} \, , \quad n \in \mathbb{N} \, , \tag{20}$$

we set

$$a_{n+1} := a_n - A_n^\sharp (F^\epsilon(a_n) - y^\delta) + P_{V_n^\perp}(a_0 - a_n) \, , \tag{21}$$

where

$$A_n^\sharp := (r_{k(n)}^n A_n)^\dagger r_{k(n)}^n \ ,$$

and

$$\begin{aligned} V_n \ := \mathcal{N}(r_{k(n)} A_n)^\perp \ &= \ \{z \in X : r_{k(n)}^n A_n a = 0 \Rightarrow \langle z, a \rangle = 0\} \\ &= \ \operatorname{span}\{A_n^* \psi_1^n, \ldots, A_n^* \psi_{k(n)}^n\} \, . \end{aligned}$$

As the convergence analysis shows, the additional term $P_{V_n^\perp}(a_0 - a_n)$, (where P_M denotes the projection operator onto the linear space M) has a stabilizing effect that also makes it possible to show convergence rates under source type conditions. Its role is comparable to the one of the additional term in the iteratively regularized Gauß-Newton method compared to the Levenberg-Marquardt method (cf. Section 11 in [8]). Note that in the case $A_n :\equiv A_0$ one easily sees by an inductive argument and condition (16) that $P_{V_n^\perp}(a_0 - a_n) = 0$ and (21) becomes

$$a_{n+1} := a_n - A_n^\sharp (F^\epsilon(a_n) - y^\delta) \, .$$

To investigate the convergence properties of method (21) we have a look at the error $e_n := a_n - a^\dagger$:

$$\begin{aligned} e_{n+1} \ &= \ -A_n^\sharp (F(a_n) - y - A_n e_n) \\ &\quad\ +A_n^\sharp (F^\epsilon(a_n) - F(a_n)) \\ &\quad\ +A_n^\sharp (y^\delta - y) \\ &\quad\ +P_{V_n^\perp}(a_0 - a^\dagger) \end{aligned}$$

It is of interest to consider not only the iteration error e_n but also its image under A_n, since $A_n e_{n+1}$ is a first order approximation for the residual $F(a_{n+1}) - F(a^\dagger)$:

$$\begin{aligned} A_n e_{n+1} \ &= \ -A_n A_n^\sharp (F(a_n) - y - A_n e_n) \\ &\quad\ +A_n A_n^\sharp (F^\epsilon(a_n) - F(a_n)) \\ &\quad\ +A_n A_n^\sharp (y^\delta - y) \\ &\quad\ +A_n P_{V_n^\perp}(a_0 - a^\dagger) \ . \end{aligned}$$

With

$$\epsilon_n := \|F^\epsilon(a_n) - F(a_n)\|$$

one obtains the recursive estimates

$$\|e_{n+1}\| \leq \|A_n^\sharp\| \begin{cases} (\frac{\eta}{1-\eta}+\frac{\|I-R\|}{1-\|I-R\|})\|A_0 e_n\| & \text{if (10), (15), and (9) holds} \\ \frac{L}{2}\|e_n\|^2 & \text{if (7), and (14) holds} \end{cases} \\ +\|A_n^\sharp\|(\epsilon_n+\delta)+\|P_{V_n^\perp}(a_0-a^\dagger)\| \ , \tag{22}$$

and

$$\|A_n e_{n+1}\| \leq \|A_n A_n^\sharp\| \begin{cases} (\frac{\eta}{1-\eta}+\frac{\|I-R\|}{1-\|I-R\|})\|A_0 e_n\| & \text{if (10), (15), and (9) holds} \\ \frac{L}{2}\|e_n\|^2 & \text{if (7), and (14) holds} \end{cases} \\ +\|A_n A_n^\sharp\|(\epsilon_n+\delta)+\|A_n P_{V_n^\perp}(a_0-a^\dagger)\| \ . \tag{23}$$

As already mentioned, $A_n e_{n+1} \approx F(a_{n+1})-y$, which should obviously become of the order of magnitude of the data error $O(\delta)$ at an appropriate stopping index N. Hence we conclude from (23) that $\|A_n A_n^\sharp\|$ should be uniformly bounded. In fact it is natural to demand this: Also for the generalized inverse $A_n^\dagger$, (which is in general unbounded in the ill–posed situation and for which the outer inverse $A_n^\sharp$ is in some sense a bounded — but not uniformly bounded — approximation,) one has $\|A_n A_n^\dagger\| \leq 1$.

The following Lemma gives conditions, under which this uniform boundedness of $\|A_n A_n^\sharp\|$ can be achieved. Moreover a bound for $\|A_n^\sharp\|$ is given, which, as gets obvious by (22), is crucial for estimating the stability of our method with respect to data errors.

Lemma 1. *Under conditions (17), (18), (20) one obtains the estimate*

$$\|A_n^\sharp\| = \sup_{v\in \mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}} \frac{\|v\|}{\|A_n^* v\|} \leq \frac{C_\psi}{c_\psi\sqrt{\lambda_{min}(M^n)}} \ , \tag{24}$$

where $\lambda_{min}(M^n)$ denotes the smallest eigenvalue of the (under our conditions nonsingular) $k(n)\times k(n)$ matrix $M^n := (\langle A_n^*\psi_i^n, A_n^*\psi_j^n\rangle_{1\leq i,j\leq k(n)})$. If, in addition

$$\sup_{\alpha\in \mathbb{R}^{k(n)}} \frac{\alpha^T S^n \alpha}{\alpha^T (M^n)^2 \alpha} \leq C_\lambda \tag{25}$$

where $S^n := [\langle A_n A_n^*\psi_i^n, A_n A_n^*\psi_j^n\rangle]_{1\leq i,j\leq k(n)}$, holds, or if

$$\mathrm{span}\{A_n A_n^*\psi_1^n,\ldots,A_n A_n^*\psi_{k(n)}^n\} \subseteq \mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\} \tag{26}$$

then

$$\|A_n A_n^\sharp\| \leq C_\psi \max\{\sqrt{C_\lambda}, 1/c_\psi\} \tag{27}$$

*Proof.*For the proof of (24) see [10].

Let now $b \in Y$ be arbitrary, $s := A_n^\sharp b$. From (13) (with $B := A_n$) one sees that $z := A_n s = A_n A_n^\sharp b$ fulfills

$$\begin{array}{ll} \langle z, \psi_j^n\rangle = \langle b, \psi_j^n\rangle & j \in \{1,\ldots,k(n)\} \\ z \in \mathrm{span}\{A_n A_n^*\psi_1^n,\ldots,A_n A_n^*\psi_{k(n)}^n\} & \end{array} ,$$

so with (18), (25) we get, with $\alpha := (\alpha_1,\ldots,\alpha_{k(n)})^T$ being the vector of coefficients of z with respect to $\{A_n A_n^*\psi_1^n,\ldots,A_n A_n^*\psi_{k(n)}^n\}$,

$$\begin{aligned} \|z\|^2 &= \alpha^T S^n \alpha \\ &\leq C_\lambda \alpha^T (M^n)^2 \alpha \\ &= C_\lambda \textstyle\sum_{j=1}^{k(n)} \langle z, \psi_j^n\rangle^2 \\ &\leq C_\psi^2 C_\lambda \|b\|^2 \ , \end{aligned}$$

If (26) holds, then, by (17), (18),

$$\|z\| \le \frac{1}{c_\psi}\sqrt{\sum_{j=1}^{k(n)}\langle z,\psi_j^n\rangle^2} = \frac{1}{c_\psi}\sqrt{\sum_{j=1}^{k(n)}\langle b,\psi_j^n\rangle^2} \le \frac{C_\psi}{c_\psi}\|b\| \,.$$

□

Note that if the A_n are compact (as typical for an ill-posed problem), then $\lambda_{min}(M^n)$ cannot be bounded away from zero as $n \to \infty$, so by the (sharp) estimate (24), $\|A_n^\sharp\|$ can grow unboundedly as $n \to \infty$.

For the convergence properties of a_n obviously the convergence of $P_{V_n^\perp}(a_0 - a^\dagger)$ (and of $A_0 P_{V_n^\perp}(a_0 - a^\dagger)$, respectively) to zero and its speed of convergence are crucial. Under conditions (16), (19), (20) it can be shown (see Section 3.3 in [8]) that in the situation of (15)

$$P_{V_n^\perp}(a_0 - a^\dagger) \to 0 \,, \qquad \text{as } n \to \infty$$

holds. To estimate the speed of convergence of $P_{V_n^\perp}(a_0 - a^\dagger)$, $A_0 P_{V_n^\perp}(a_0 - a^\dagger)$ to zero, one can use the smoothing properties of the operators A_n, A_n^* (see the remarks below).

Remark 1. From the linear theory it is known that the choice $\psi_j^n := v_j^n$, where $(\sigma_j^n; u_j^n, v_j^n)_{j\in\mathbb{N}}$ is a singular system of A_n, i.e., truncated singular value expansion, is optimal not only from the point of view of convergence of $P_{V_n^\perp}(a_0 - a^\dagger)$ and $A_n P_{V_n^\perp}(a_0 - a^\dagger)$ but also for stability reasons. One has, for any basis $\{\psi_1^n, \ldots, \psi_{k(n)}^n\}$ that under a source condition (5)

$$\|P_{\mathrm{span}\{A_n^*\psi_1^n,\ldots,A_n^*\psi_{k(n)}^n\}^\perp}(a_0 - a^\dagger)\| \ge (\sigma_{k(n)+1}^n)^{2\nu} = \|P_{\mathrm{span}\{A_n^* v_1^n,\ldots,A_n^* v_{k(n)}^n\}^\perp}(a_0 - a^\dagger)\|$$

and that

$$\lambda_{\min}(\langle A_n^*\psi_i^n, A_n^*\psi_j^n\rangle_{1\le i,j\le k(n)}) \le \sigma_{k(n)}^n = \lambda_{\min}(\langle A_n^* v_i^n, A_n^* v_j^n\rangle_{1\le i,j\le k(n)}) \,.$$

Moreover (26) holds and $\|A_n A_n^\sharp\| = 1$ if one sets $\psi_j^n := v_j^n$. Note that due to the fact that

$$\frac{\lambda_{\max}([\langle A_n A_n^* v_i^n, A_n A_n^* v_j^n\rangle]_{1\le i,j\le k(n)})}{(\lambda_{\min}([\langle A_n^* v_i^n, A_n^* v_j^n\rangle]_{1\le i,j\le k(n)}))^2} = \left(\frac{\sigma_1^n}{\sigma_{k(n)}^n}\right)^4 \to \infty \text{ as } k(n) \to \infty \,,$$

(25) is not a typical condition in the usual analysis of truncated SVE.

Due to the often large effort for calculating a singular system, the choice $\psi_j^n := v_j^n$ only makes sense in the context of a frozen Newton method, i.e., if (10) and (15) holds together with (9). One then obtains local convergence of (21) with the rate

$$\|a_n - a^\dagger\| = O((\sigma_{k(n)+1}^n)^{2\nu})$$

under a source condionion of the type (5) in the noiseless case $\delta = \epsilon_n = 0$; in the case of noisy data one gets, with an appropriate (a priori or a posteriori) choice of the stopping index N and of the accuracy ϵ_n,

$$\|a_N - a^\dagger\| = O(\delta^{\frac{2\nu}{2\nu+1}})$$

as $\delta \to 0$ (cf. [2], [19], [20]).

In this paper, however, we concentrate on a completely different type of basis functions ψ_j^n, namely on finite element bases. Here it turns out that (25) is a natural condition, and therefore $\|A_n A_n^\sharp\|$ is uniformly bounded. For estimating the convergence rates of $A_0 P_{V_n^\perp}(a_0 - a^\dagger)$, and of $P_{V_n^\perp}(a_0 - a^\dagger)$ under a source condition, the smoothing properties of the operators A_0, A_0^* can be used in the context of parameter identification in a PDE, as follows: Assume that X is some

Sobolev space H^s and $Y = H^t$ for some $s, t \in \mathbb{R}$ and that A_n maps not only functions in H^s but even H^{s-p}-functions into Y where p is some positive number quantifying the smoothing that is done by A_n; let, moreover A_n^* be a continuous operator from H^{t-p-q} to H^{s-p}, and $\mathcal{N}(A_n)^\perp \subseteq A_n^* H^{t-q}$, with $q > 0$ being the degree of smoothing of A_n^*. Then there exists a $w \in H^{t-q}$ such that $a_0 - a^\dagger = A_n^* w$ and

$$\begin{aligned} \|A_n P_{V_n^\perp}(a_0 - a^\dagger)\|_Y &\leq \|A_n\|_{H^{s-p}\to H^t} \|P_{V_n^\perp} A_n^* w\|_{H^{s-p}} \\ &= \|A_n\|_{H^{s-p}\to H^t} \inf_{v \in \mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}} \|A_n^* v - A_n^* w\|_{H^{s-p}} \\ &\leq \|A_n\|_{H^{s-p}\to H^t} \|A_n^*\|_{H^{t-p-q}\to H^{s-p}} \|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{H^{t-p-q}} \ . \end{aligned}$$

If the ψ_j^n are finite elements with mesh size h_n of order not less than $t-q-1$ then the term $\|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{H^{t-p-q}}$ can be expected to go to zero with the rate $O((h_n)^p)$ (cf. Theorem 4.1 in [14]).

Similarly, one obtains an estimate of the speed of convergence of $P_{V_n^\perp}(a_0 - a^\dagger)$ in the situation of (5) with $\nu = \frac{1}{2}$ i.e, an estimate of $P_{V_n^\perp} F'(a^\dagger)^* w$: If $X = H^s$, $Y = H^t$ and A_n^* maps H^{t-q} continuously into H^s, (where $q > 0$), then

$$\begin{aligned} \|P_{V_n^\perp}(a_0 - a^\dagger)\|_X &= \|P_{V_n^\perp} F'(a^\dagger)^* w\|_X \\ &\leq \|A_n^* - F'(a^\dagger)^*\|_{Y\to X} \|w\|_Y + \|P_{V_n^\perp} A_n^* w\|_X \\ &\leq L\|e_n\|_X \|w\|_Y + \|A_n^*\|_{H^{t-q}\to H^s} \|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{H^{t-q}} \ , \end{aligned}$$

if (7), (14) holds, or

$$\begin{aligned} \|P_{V_n^\perp}(a_0 - a^\dagger)\|_X &= \|P_{V_n^\perp} A_0^* (R^{-1})^* w\|_X \\ &\leq \|A_0^*\|_{H^{t-q}\to H^s} \|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} (R^{-1})^* w\|_{H^{t-q}} \ , \end{aligned}$$

if (10), (15) holds, respectively, where $w \in Y$. Again, using finite elements of order $t-1$, one can expect that $\|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{H^{t-q}} = O((h_n)^q)$. Due to the fact that if $\|A_n\|_{H^{s-p}\to H^t} \leq C_A$ for all $s, t \in \mathbb{R}$ then also $\|A_n^*\|_{H^{t-p}\to H^s} \leq C_A$ for all $s, t \in \mathbb{R}$ (here * denotes the adjoint with respect to the H^s- and the H^t-topology) the two exponents p and q are likely to be the same.

By these smoothness properties and the fact that $p = q$, also (25) can be established and one obtains an estimate for $\lambda_{\min}(M^n)$. (cf. Section 3 below).

The following theorem summarizes the considerations of this section and states some convergence and convergence rates results for (21).

Theorem 1. *Let the sequence a_n be defined by (21), where the basis functions ψ_j^n are finite elements on a grid of size h_n that fulfill assumptions (16), (17), (18), (19), (20) and*

$$h_n \to 0 \ , \quad \text{as } n \to \infty \ , \qquad 1 \leq \frac{h_n}{h^{n+1}} \leq r \tag{28}$$

for some $r > 1$. Let, moreover (25) (or (26)) hold and let $\kappa \in \mathbb{R}^+$ be such that, for some constant $c_\lambda > 0$

$$\lambda_{\min}(M^n) \geq c_\lambda (h_n)^\kappa \tag{29}$$

1. *Let (10), (11), (9), (15) with $\|a_0 - a^\dagger\|$, $\|I - R\|$, η sufficiently small hold and assume that*

$$\|A_0 P_{V_n^\perp}(a_0 - a^\dagger)\| = o((h_n)^{\frac{\kappa}{2}}) \tag{30}$$

and that the approximation accuracy of $F^\epsilon(a_n)$ for $F(a_n)$ is chosen such that

$$\epsilon_n = o((h_n)^{\frac{\kappa}{2}}) \ . \tag{31}$$

Then, in the noiseless case $\delta = 0$ the iterates converge as $n \to \infty$:

$$a_n \to a^\dagger \ , \quad \text{as } n \to \infty \ .$$

In case of noisy data and if the iteration is stopped at an index $N = N(\delta)$ with

$$N \to \infty \ , \quad \text{and } \frac{\delta^2}{(h_n)^\kappa} \to 0 \ , \quad \text{as } \delta \to 0 \tag{32}$$

one obtains

$$a_N \to a^\dagger \ , \quad \text{as } \delta \to 0 \ . \tag{33}$$

If additionally $P_{V_n^\perp}(a_0 - a^\dagger)$ not only goes to zero but converges at some rate

$$\|P_{V_n^\perp}(a_0 - a^\dagger)\| = O((h_n)^\mu) \tag{34}$$

for some $\mu > 0$ and if

$$\|A_0 P_{V_n^\perp}(a_0 - a^\dagger)\| = O((h_n)^{\tilde{\mu}+\frac{\kappa}{2}}) \ , \tag{35}$$

for some $\tilde{\mu}$, which is nonnegative, due to (30), as well as

$$\epsilon_n = O((h_n)^{\min\{\mu,\tilde{\mu}\}+\frac{\kappa}{2}}) \ , \tag{36}$$

then, if $\delta = 0$, a_n converges at the rate

$$\|a_n - a^\dagger\| = O((h_n)^{\kappa \cdot \underline{\nu}}) \tag{37}$$

with

$$\underline{\nu} := \frac{\min\{\mu, \tilde{\mu}\}}{\kappa} \ .$$

If $\delta \neq 0$ and N is chosen such that

$$(h_n)^{\kappa \cdot (\underline{\nu}+\frac{1}{2})} \le \tau\delta < (h_n)^{\kappa \cdot (\underline{\nu}+\frac{1}{2})} \quad \forall n \le N-1 \tag{38}$$

for some fixed and sufficiently large $\tau > 0$, then

$$\|a_N - a^\dagger\| = O(\delta^{\frac{2\underline{\nu}}{2\underline{\nu}+1}}) \ .$$

If, in addition to (35) one has

$$\|A_0 P_{V_n^\perp}(a_0 - a^\dagger)\| \ge \underline{c}((h_n)^{\bar{\mu}+\frac{\kappa}{2}}) \ ,$$

with some constant $\underline{c} > 0$ and some $\bar{\mu} \ge \tilde{\mu}$, then, using the a posteriori parameter choice

$$\epsilon_n \le c_\epsilon \|F^\epsilon(a_n) - y^\delta\| \tag{39}$$

$$\begin{aligned} &\max\{\|F^\epsilon(a_{N-1}) - y^\delta\|, \|A_0 A_n^\sharp(F^\epsilon(a_{N-1}) - y^\delta + A_0(a_0 - a_{N-1}))\| \le \tilde{\tau}\delta \\ &< \max\{\|F^\epsilon(a_{n-1}) - y^\delta\|, \|A_0 A_n^\sharp(F^\epsilon(a_{n-1}) - y^\delta + A_0(a_0 - a_{n-1}))\| \quad \forall n \le N-1 \end{aligned} \tag{40}$$

with some sufficiently small $c_\epsilon > 0$ and some sufficiently large $\tilde{\tau} > 0$, instead of the a priori choice (31), (32), (36), (38), one gets convergence (33) and, if (34) holds,

$$\|a_N - a^\dagger\| = O(\delta^{\frac{2\underline{\nu}}{2\overline{\nu}+1}}) \ , \tag{41}$$

where $\overline{\nu} := \frac{\bar{\mu}}{\kappa}$.

2. *Let (7), (14) with $\|a_0 - a^\dagger\|$ sufficiently small hold and assume that*

$$\|P_{V_n^\perp}(a_0 - a^\dagger)\| \le c_P (h_n)^\mu \tag{42}$$

with

$$\mu \ge \frac{\kappa}{2}$$

and that the approximation accuracy of $F^\epsilon(a_n)$ for $F(a_n)$ is chosen such that

$$\epsilon_n \leq c_\epsilon (h_n)^{\mu+\frac{\kappa}{2}} \ , \tag{43}$$

where c_P and c_ϵ are sufficiently small constants.
Then if $\delta = 0$ one obtains

$$\|a_n - a^\dagger\| = O((h_n)^{\kappa\cdot\underline{\nu}}) \ , \tag{44}$$

with

$$\underline{\nu} := \frac{\mu}{\kappa} \ .$$

If $\delta \neq 0$ and N is chosen according to (38) for some fixed and sufficiently large $\tau > 0$, then

$$\|a_N - a^\dagger\| = O(\delta^{\frac{2\underline{\nu}}{2\underline{\nu}+1}}) \ .$$

*Proof.*Consider first case 1.: By our assumptions and Lemma 1 we obtain

$$\begin{aligned} \frac{\|A_0 e_{n+1}\|}{(h^{n+1})^{\frac{\kappa}{2}}} &\leq r^{\frac{\kappa}{2}} c \frac{\|A_0 e_n\|}{(h^n)^{\frac{\kappa}{2}}} + r^{\frac{\kappa}{2}} C_\psi \max\{\sqrt{C_\lambda}, 1/c_\psi\} \frac{\epsilon_n + \delta}{(h^n)^{\frac{\kappa}{2}}} + r^{\frac{\kappa}{2}} \frac{\|A_0 P_{V_n^\perp}(a_0 - a^\dagger)\|}{(h^n)^{\frac{\kappa}{2}}} \\ &\leq (r^{\frac{\kappa}{2}} c)^{n+1} \frac{\|A_0 e_0\|}{(h^0)^{\frac{\kappa}{2}}} + \sum_{j=0}^{n} (r^{\frac{\kappa}{2}} c)^{n-j} \gamma_j \end{aligned} \tag{45}$$

with

$$c := (\frac{\eta}{1-\eta} + \frac{\|I - R\|}{1 - \|I - R\|}) C_\psi \max\{\sqrt{C_\lambda}, 1/c_\psi\} \ ,$$

which can be made smaller that $r^{-\frac{\kappa}{2}}$ by making $\|I - R\|$, η sufficiently small, and

$$\gamma_j := r^{\frac{\kappa}{2}} \frac{C_\psi \max\{\sqrt{C_\lambda}, 1/c_\psi\}(\epsilon_j + \delta) + \|A_0 P_{V_j^\perp}(a_0 - a^\dagger)\|}{(h^j)^{\frac{\kappa}{2}}} \ .$$

So since $r^{\frac{\kappa}{2}} c < 1$ and $\gamma_j \to 0$ as $j \to \infty$ we obtain that

$$\frac{\|A_0 e_n\|}{(h_n)^{\frac{\kappa}{2}}} \to 0 \text{ as } n \to \infty \ .$$

in the noiseless case $\delta = 0$. Inserting this into (22):

$$\|e_{n+1}\| \leq \frac{C_\psi}{c_\psi c_\lambda} (\frac{\eta}{1-\eta} + \frac{\|I - R\|}{1 - \|I - R\|}) \frac{\|A_0 e_n\|}{(h_n)^{\frac{\kappa}{2}}} + \frac{C_\psi}{c_\psi c_\lambda} \frac{\epsilon_n + \delta}{(h_n)^{\frac{\kappa}{2}}} + \|P_{V_n^\perp}(a_0 - a^\dagger)\| \tag{46}$$

yields convergence $a_n \to a^\dagger$ as $n \to \infty$ if $\delta = 0$. In the case of nonvanishing noise $\delta \neq 0$ convergence $a_N \to a^\dagger$ as $\delta \to 0$ is obtained by using (32) in (45), (46). The convergence rates results can be derived similarly: One gets

$$\begin{aligned} \|A_0 e_n\| &\leq C(h_n)^{\kappa\cdot(\underline{\nu}+\frac{1}{2})} \\ \|e_n\| &\leq C(h_n)^{\kappa\underline{\nu}} \end{aligned} \qquad \forall n \leq N (\text{or } \forall n \in \mathbb{N} \text{ if } \delta = 0)$$

with some constant $C > 0$ independent of n and δ. Therefore we obtain (37) if $\delta = 0$ and, in the case of noisy data, by (38):

$$\|e_N\| \leq C(h_n)^{\kappa\underline{\nu}} \leq C \tau^{\frac{\underline{\nu}}{\underline{\nu}+\frac{1}{2}}} \delta^{\frac{\underline{\nu}}{\underline{\nu}+\frac{1}{2}}} \ .$$

In the Lipschitz case 2. we proceed analogously, using the estimate

$$\frac{\|e_{n+1}\|}{(h^{n+1})^\mu} \leq \frac{C_\psi L r^\mu}{2 c_\psi c_\lambda} \Big(\frac{\|e_n\|}{(h^n)^\mu}\Big)^2 + \frac{C_\psi r^\mu}{c_\psi c_\lambda} \frac{\epsilon_n + \delta}{(h_n)^{\mu+\frac{\kappa}{2}}} + r^\mu \frac{\|P_{V_n^\perp}(a_0 - a^\dagger)\|}{(h^n)^\mu}$$

that results from (22) and that, by (42), (43), (38), and if $\frac{\|e_0\|}{(h^0)^\mu}$ is sufficiently small, inductively yields uniform boundedness of $\frac{\|e_n\|}{(h^n)^\mu}$ for $n \leq N(\delta)$ and all δ in case of noisy data or for all $n \in \mathbb{N}$ if $\delta = 0$.

The a posteriori parameter choice (40) in case 1. results from carrying over the idea of the a posteriori parameter chiose proposed in [20] to method (21). Its convergence proof can be done along the lines of the proofs of Lemma 2.1, Theorem 2.3, and Theorem 2.4 in [20], with $r_{\alpha_n}(A_n^* A_n)(a_0 - a^\dagger)$ in [20] replaced by our $P_{V_n^\perp}(a_0 - a^\dagger)$, $g_{\alpha_n}(A_n^* A_n)A_n$ replaced by $A_n^\sharp$, and α_n replaced by $(h_n)^\kappa$. Moreover one has to take into account that by (39), (9) and (10)

$$\epsilon_n \leq \frac{c_\epsilon}{1-c_\epsilon}\left(\frac{1}{(1-\eta)(1-\|I-R\|)}\|A_0 e_n\| \;+\; \delta\right)$$

holds. □

As gets obvious by Proposition 1, the mesh size h_n of the finite elements plays the role of a regularization parameter. The quantity $\underline{\nu}$ corresponds to the smoothnmess parameter ν in (5) in the sense that, by the results cited in remark 1, one has $\underline{\nu} \leq \nu$.

The a posteriori parameter choice (39), (40) has the advantage that one obtains the optimal rate (41) without needing any explicite knowledge on the exponents $\mu, \tilde{\mu}, \bar{\mu}$.

From (32) and (38) it follows that

$$\delta \leq C(h^{N-1})^{\frac{\kappa}{2}}$$

for some $C > 0$ independent of δ. If the sequence of mesh sizes is chosen as fast decreasing as possible, subject to the constraint (28), i.e., if

$$h_n = O(r^{-n}) \;,$$

this yields

$$N(\delta) = O(|\log(\delta)|)$$

so the number of Newton steps grows only logarithmically as $\delta \to 0$.

Since the finite-dimensional linear system (13) (with B, b replaced by $A_n, (F^\epsilon(a_n) - y^\delta)$, respectively) that has to be solved in each Newton step of (21) comes from a finite-element discretization in the setting of proposition 1, it is near at hand to apply a multigrid method to it. However, multigrid methods behave completely different for ill–posed problems (12) than they do for well–posed ones. Only recently some first investigations on multigrid methods for ill-posed problems have been made (cf. [4], [18], [21]).

Remark 2. If conditions (9), (10) in the assumptions of the first part of Proposition 1 are replaced by

$$F'(\bar{a}) = F'(a)R_a^{\bar{a}} \qquad \bar{a}, a \in \mathcal{B}_{2\rho}(a_0) \tag{47}$$

$$A_0 = F'(a^\dagger)R$$

for some bounded linear operators $R_a^{\bar{a}}$, A_0, R with $\|R_a^{\bar{a}} - I\| \leq C_R\|\bar{a} - a\|$, $\|I - R\| < 1$ (and sufficiently small), then the assertions of Proposition 1.1 remain valid, since we have

$$\begin{array}{rcl} e_{n+1} & = & -A_n^\sharp A_0(R^{-1}\int_0^1 R_{a^\dagger}^{a^\dagger+\theta(a_n-a^\dagger)} - I)e_n \\ & & +A_n^\sharp(F^\epsilon(a_n) - F(a_n)) \\ & & +A_n^\sharp(y^\delta - y) \\ & & +P_{V_n^\perp}(a_0 - a^\dagger) \;, \end{array}$$

where

$$\|A_n^\sharp A_0\| \;=\; \|P_{V_n}\| \;=\; 1$$

and

$$\|R^{-1}\int_0^1 R_{a^\dagger}^{a^\dagger+\theta(a_n-a^\dagger)} - I\| \le \frac{2\rho C_R + \|I-R\|}{1-\|I-R\|} \ .$$

Note that in this situation no such condition as (25) (or (26)) is needed. The range invariance condition (47), that is by the way closely related to the so-called affine covariant Lipschitz condition in [6], holds for several parameter identification problems, especially for identification from boundary measurements, where (8) or (9) can sometimes not be verified.

Remark 3. For a convergence analysis of (21) with general bases ψ_j^n see [2]. Since $A_n^\sharp$ is an outer inverse of A_n, i.e., $A_n^\sharp A_n A_n^\sharp = A_n^\sharp$ holds, (21) also fits into the more general framework of Newton type methods using outer inverses, ([24], see also [2]).

3 A model problem

In order to demonstrate the ideas of our convergence proof we concentrate on the one-dimensional case in (1)

$$\begin{array}{c} -(a\cdot u_x)_x = 0 \quad \text{in } (0,1) \\ u(0) = g^0 \ , \quad u(1) = g^1 \ . \end{array} \tag{48}$$

Here the forward operator F evaluated at a parameter a is defined by the (weak) solution u of (48):

$$\begin{array}{rcl} F:\mathcal{D}(F) & \to & Y \\ a & \mapsto & u \end{array}$$

where $\mathcal{D}(F) := \{a \in H^1(0,1) : a \ge \underline{a} > 0 \ , \ a(0)=a^0 \ , \ a(1)=a^1\}$ (with two fixed known values a^0, a^1) and $g^1 \ne g^0$, so that by (48) u_x is bounded away from zero. Note that one naturally has measurements u^δ that approximate the exact data u only in the L^2-sense

$$\|u^\delta - u\|_{L^2} \le \delta \ ,$$

so although $u \in H^2(0,1)$ as soon as $a \in \mathcal{D}(F)$, one takes $Y = L^2(0,1)$. It can be shown (see [9]) that the Frechet derivative of F and its adjoint are given by

$$\begin{array}{rcll} F'(a)b & = & A(a)^{-1}[(b\cdot F(a)_x)_x] & b \in H_0^1(0,1) \\ F'(a)^*w & = & -A(1)^{-1}[F(a)_x \cdot (A(a)^{-1}w)_x] & w \in L^2(0,1) \ , \end{array}$$

where

$$\begin{array}{rcl} A(a): H^2(0,1)\cap H_0^1(0,1) & \to & L^2(0,1) \\ v \mapsto -(a\cdot v_x)_x \ , & & \end{array}$$

$H_0^1(0,1)$ is equipped with the inner product

$$\langle b^1, b^2\rangle_{H_0^1} := \int_0^1 b_x^1(x)\cdot b_x^2(x)\, dx \ ,$$

and we use the fact that $A(1)^{-1}$ is the adjoint of the embedding operator from $H^2(0,1)\cap H_0^1(0,1)$ into $L^2(0,1)$.

One can show, (cf. [17]) that with

$$G(a)^*w := -A(1)^{-1}[u_x^\delta \cdot (A(a)^{-1}w)_x]$$

one obtains

$$G(\bar a) = R_a^{\bar a} G(a) + E_a^{\bar a}$$

for some bounded linear operators $R_a^{\bar a}$, $E_a^{\bar a}$ with

$$\|I - R_a^{\bar a}\| < C_R\|a-\bar a\| \ , \quad \|E_{a^\dagger}^a\| \le C_E\delta \ ,$$

whence

$$\|F(a) - F(a^\dagger) - G(a^\dagger)(a - a^\dagger)\| \le C_T \Big(\|a - a^\dagger\| \, \|F(a) - F(a^\dagger)\| + \delta \Big) \qquad a \in \mathcal{B}_{2\rho}(a_0) \ ,$$

where C_R, C_E, C_T are some constants independent of $a, \bar{a}$, and δ. An inspection of the proof of Proposition 1 shows that its assertion still holds if $F'(a)$ is replaced by $G(a)$. If in the definition of $\mathcal{D}(F)$ $a^1 = a^0$ (which we will assume in the following for reasons of simplicity) one can choose the starting value a_0 to be just the constant function with value a^0 and the operator A_0 as the Frechet derivative of F at a_0:

$$A_0 b := F'(a_0) b = \frac{g^1 - g^0}{a^0} A(1)^{-1} [b_x] \tag{49}$$

for which (10) holds with $R^{-1\ *} w := \frac{a^0}{g^1 - g^0} \Big(F(a^\dagger)_x \cdot (P_{[F(a^\dagger)]^\perp} A(a^\dagger)^{-1} w)_x \Big)_x$ (here P denotes the projection with respect to the H_0^1 inner product).

It can be shown (see [9]) that for our parameter identification problem the source condition $a_0 - a^\dagger = F'(a^\dagger)^* w$ for some sufficiently small $w \in L^2(0,1)$ i.e., (5) with $\nu = \frac{1}{2}$, is equivalent to

$$a_0 - a^\dagger \in H^3(0,1) \cap H_0^1(0,1) \quad \wedge \quad \int_0^1 \frac{A(1)(a_0 - a^\dagger)}{(F(a^\dagger))_x} \, dx = 0 \tag{50}$$

with $\|A(1)(a_0 - a^\dagger)\|_{H^1(0,1)}$ being sufficiently small. So, like in many other inverse problems, the source condition is some smoothness condition on the difference between the initial guess and the exact solution, together with some boundary conditions. Taking the closure of $\mathcal{R}(F'(a^\dagger)^*)$ with respect to the H_0^1-norm, one obtains the nullspace of $F'(a^\dagger)$:

$$\mathcal{N}(F'(a^\dagger))^\perp = \{ b \in H_0^1(0,1) \ : \ (A(1)b)(\frac{1}{F(a)_x}) = 0 \} \ . \tag{51}$$

Here $A(1)$ is extended to $H_0^1(0,1)$ by using the weak formulation of the derivative:

$$\begin{array}{rcl} A(1) : H_0^1(0,1) & \to & (H^1(0,1))^* \\ v & \mapsto & \left\{ \begin{array}{rcl} A(1)v : H^1(0,1) & \to & \mathbb{R} \\ \phi & \mapsto & \int_0^1 v_x \cdot \phi_x \, dx \end{array} \right. \end{array} .$$

The smoothing properties of the operators $F'(a)$, $F'(a)^*$ can be quantified as follows. (In the following $\overline{M}^H$ denotes the closure of the set M with respect to the topology of the Banach space H.)

Lemma 2. *For any $a \in \mathcal{D}(F)$, $F'(a)$ can be continuously extended to an operator from $\overline{H_0^1}^{(H^1(0,1))^*}$ to $L^2(0,1)$.*

*Proof.*See the Appendix.

Lemma 3. *For any $a \in \mathcal{D}(F)$ with $F(a) \in H^4(0,1)$, there holds*

$$\forall b \in \mathcal{N}(F'(a^\dagger))^\perp \ \exists w \in (H^2(0,1))^* \ : b = F'(a)^* w \tag{52}$$

and $F'(a)^*$ can be continuously extended to an operator from $\overline{L^2(0,1)}^{(H^4(0,1))^*}$ to $\overline{H_0^1(0,1)}^{(H^1(0,1))^*}$.

*Proof.*See the Appendix.

So in terms of the remarks in Section 2 we have $p = q = 2$ for the smoothness parameters p and q, respectively.

A discretization of the forward problem of evaluating $F(a_n)$ by finite elements could be done as follows: We divide the unit interval by $\frac{1}{H_n}-1$ equidistant points into $\frac{1}{H_n}$ intervals of size H_n. Inserting ansatz and test functions ϕ_j $(1 \le j \le K(n))$ into the weak formulation of (48) yields the finite dimensional problem

$$\begin{array}{l} (F(a_n))(x) = g^0 + (g^1 - g^0)x + \sum_{i=1}^{k(n)} \alpha_i \phi_i \\ \forall j \in \{1,\ldots,K(n)\} \ : \quad \sum_{i=1}^{K(n)} \alpha_i \int_0^1 a_n(x)\phi_{ix}(x)\phi_{jx}(x)\, dx = (g^1 - g^0)\int_0^1 a_n(x)\phi_{jx}(x)\, dx \ , \end{array}$$

a system of $K(n)$ linear algebraic equations in $K(n)$ variables, whose (approximate) solution we obtain by some linear solver — here a multigrid method suggests itself. For a PDE of second order it suffices to use finite elements of first order, which are, in our one-dimensional situation, the hat functions

$$\phi_j(x) := \sqrt{\frac{3}{2h^3}}\Big(\chi_{[(j-1)h,jh]}(x)(x-(j-1)h) + \chi_{[jh,(j+1)h]}(x)((j+1)h - x)\Big) \tag{53}$$

with $h := H_n$. By the convergence theory of finite elements (cf. Theorem 4.3 in [14]) this yields an approximation $F^\epsilon(a_n)$ of $F(a_n)$ with

$$\epsilon_n := \|F^\epsilon(a_n) - F(a_n)\|_{L^2(0,1)} = C \cdot (H_n)^2$$

where (due to the fact that $\|a_n\|_{H^1}$ and therefore also $\|F(a_n)\|_{H^2}$ are uniformly bounded) $C > 0$ can be chosen to be a constant independent of n.

Method (21) for solving the inverse problem of identifying a in (48) includes the possibilty to make use of finite elements also as a basis ψ_j^n for the finite-dimensional subspaces onto that the linearized equation is projected in each Newton step. (cf. Proposition 1). In order to obtain the expected convergence rate $O((h_n)^p)$ of the term $\|P_{\mathrm{span}\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{(H^{l+p})^*}$ with $w \in (H^l(0,1))^*$, (cf. the corresending remarks in Section 2) in the error decomposition (22), (23), we need an analogue of Theorem 4.1 in [14] for Sobolev spaces of "negative" order, which we will prove similarly to this theorem with some appropriate changes:

Lemma 4. *Let the basis functions ψ_j^n be finite elements of polynomial order $l-1$ on a grid of size h_n and let $w \in (H^l(0,1))^*$. Then*

$$\|P_{span\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{(H^{l+p}(0,1))^*} \le C(h_n)^p$$

holds with some constant C independent of h_n. If, in addition, (16) and (19) hold, then

$$\frac{\|P_{span\{\psi_1^n,\ldots,\psi_{k(n)}^n\}^\perp} w\|_{(H^{l+p}(0,1))^*}}{(h_n)^p} \to 0 \ , \qquad \text{as } h_n \to 0.$$

Proof. See the Appendix.

The results of Lemmas 3, 4 suggest to use piecewise linear finite elements, i.e., $\psi_j^n := \phi_j$ as in (53) with $h := h_n$. If we do the refinement of the grid in each Newton step by bisection, i.e., choose $h_n := \frac{1}{k(n)+1}$ with $k(n) := 2^{l(n)} - 1$, where $l(n)$ is a monotonically increasing sequence of positive integers then obviously (16) holds and it can be easily checked that we also obtain (17), (18) with

$$c_\psi = \sqrt{\lambda_{\min}(\langle\psi_i,\psi_j\rangle_{1\le i,j\le k(n)})} \ge \sqrt{\frac{1}{2}} \ , \quad C_\psi = \sqrt{\lambda_{\max}(\langle\psi_i,\psi_j\rangle_{1\le i,j\le k(n)})} \le \sqrt{\frac{3}{2}} \ ,$$

by Gershgorin's theorem.

For our simple choice $A_0 := F'(a_0)$ with $a_0 :\equiv a^0$ we can give an explicit expression for $\langle A_0^*\psi_i^n, A_0^*\psi_j^n\rangle$, which, by (24) yields an estimate of $\|A_n^\sharp\|$, as well as the verification of condition (25):

Lemma 5.

$$\|A_n^\sharp\| = O((h_n)^{-2}) \ ,$$

and $\|A_n A_n^\sharp\|$ is uniformly bounded.

*Proof.*See the Appendix.

Now we can apply proposition 1 to obtain

Corollary 1. *Let a_n be defined by (21) with $a_0 :\equiv a^0$ and $A_n :\equiv A_0 := F'(a_0)$, ψ_j^n being piecewise linear finite elements (53) on a grid of size $h_n := 2^{-l(n)}$, where $l(n) \nearrow \infty$ as $n \to \infty$, and let $F^\epsilon(a_n)$ be the approximation of $F(a_n)$ by linear finite elements on a grid of size H_n with*

$$H_n = o(h_n) \ .$$

Assume that $\|a_0 - a^\dagger\|$ is sufficiently small and that $a_0 - a^\dagger \in \mathcal{N}(F'(a^\dagger))^\perp$.

Then, in the noiseless case $\delta = 0$

$$a_n \to a^\dagger \ , \ \text{as } n \to \infty \ .$$

In the case of noisy data with (3) and with the a priori choice (32) of N with $\kappa := 4$ one obtains

$$a_N \to a^\dagger \ , \ \text{as } \delta \to 0 \ . \tag{54}$$

If, in addition $a_0 - a^\dagger$ fulfills the source condition (50) and if

$$H_n = O((h_n)^2)$$

then

$$\|a_n - a^\dagger\| = O((h_n)^2) \ , \ \text{as } n \to \infty$$

if $\delta = 0$ and, if $\delta \neq 0$ and N is chosen according to (38) with $\mu = \tilde{\mu} = 2$, then

$$\|a_N - a^\dagger\| = O(\sqrt{\delta}) \ , \ \text{as } \delta \to 0 \ . \tag{55}$$

Convergence (54) and the convergence rate (55) (under condition (50)) are also obtained with the a posteriori parameter choice (39), (40) with sufficiently small c_ϵ and sufficiently large $\tilde{\tau}$.

4 Numerical Experiments

To numerically verify the theoretical results of this paper, we implemented the proposed algorithm for the model problem of Section 3 in a Matlab program.

We set $g^0 = 0$, $g^1 = 1$, $a_0 :\equiv a^0 = a^1 = 1$ and used uniform grids on $(0,1)$, with $h_n = 2^{[(n+1)/2]}$, $H_n = h_n^2$.

The calculations were done for three examples:

(ex1)
$$a^\dagger(x) = \begin{cases} 1 + \sin(\frac{\pi}{2}(\frac{x-0.5}{0.45} + 1)) & x \in [0.05, 0.95] \\ 1 & \text{else} \end{cases}$$

$$u^\dagger(x) = \begin{cases} \frac{x}{1-0.45(2-\frac{4}{\pi})} & x \in [0, 0.05] \\ \frac{(0.05+\frac{0.9}{\pi}(\tan(\frac{\pi}{4}\frac{x-0.5}{0.45})+1)}{1-0.45(2-\frac{4}{\pi})} & x \in [0.05, 0.95] \\ \frac{x-0.45(2-\frac{4}{\pi})}{1-0.45(2-\frac{4}{\pi})} & x \in [0.95, 1] \end{cases}$$

(ex2)
$$a^\dagger(x) = \begin{cases} 1 + \frac{1}{3}\sin^2(\pi\frac{x-0.5}{0.2}) & x \in [0.3, 0.7] \\ 1 & \text{else} \end{cases}$$

$$u^\dagger(x) = \begin{cases} \frac{x}{1-0.2(2-\sqrt{3})} & x \in [0, 0.3] \\ \frac{0.3+\frac{0.2\sqrt{3}}{2\pi}(\arctan(\sqrt{3}\tan(\frac{\pi}{2}\frac{x-0.5}{0.2}))+\arctan(\frac{1}{\sqrt{3}}\tan(\frac{\pi}{2}\frac{x-0.5}{0.2}))+\pi)}{1-0.2(2-\sqrt{3})} & x \in [0.3, 0.7] \\ \frac{x-0.2(2-\sqrt{3})}{1-0.2(2-\sqrt{3})} & x \in [0.7, 1] \end{cases}$$

$$\text{(ex3)}\qquad \begin{array}{l} a^\dagger(x) = 1 + \frac{1}{3}\sin^2(\pi x) \\ u^\dagger(x) = 1 + \frac{1}{\pi}(\arctan(\sqrt{3}\cot(-\frac{\pi}{2}x)) + \arctan(\frac{1}{\sqrt{3}}\cot(-\frac{\pi}{2}x)) \end{array}$$

In the third example, the source condition (5) with $\nu = \frac{1}{2}$ is fulfilled; in (ex2) we have $a_0 - a^\dagger \in H^2 \cap H_0^1$ and $\int_0^1 \frac{A(1)(a_0-a^\dagger)}{(F(a_0))_x}\,dx = 0$, so one can expect to obtain results according to $\nu = \frac{1}{4}$; in the first example no source condition is satisfied, since $a_0 - a^\dagger$ has a jump in the first derivative.

The first test was done with exact data and the predicted convergence rates $\|a_n - a^\dagger\| = o(2^{-\nu n})$ could be observed (see Table 1).

Then random noise with a noise level between 1 and 16 per cent was added in a series of five tests for each example and each noise level, in order to verify the rate $\|a_N - a^\dagger\| = O(\delta^{2\nu/(2\nu+1)})$ Here the stopping index N was chosen according to the a posteriori criterion (40) with $\tilde{\tau} := 1.1$. In Fig. 1 we plotted the values $\frac{\|a_N - a^\dagger\|}{\delta^{2\nu/(2\nu+1)}}$ over the (relative) noise level for each of the five tests. The graphics show that in each example these quotients are in fact about the same size for the different values of δ; they even show a tendency to get smaller as δ goes to zero.

Finally, (see Fig. 2,) we compared the iterates of the projection-regularized Newton method with those of Newton's method without regularization, for example (ex2) with noisy data ($\delta = 0.001$). Due to the instability of the underlying problem, the propagated data error dominates the result of Newton's method without regularization already after few iterations (dotted line), while with regularization by projection one obtains a reasonnable approximation (dashed line) to the exact parameter a (solid line).

Table 1

(ex1): $\nu = 0$		(ex2): $\nu = \frac{1}{4}$		(ex3): $\nu = \frac{1}{2}$	
n	$\frac{\|a_n - a^\dagger\|}{2^{-\nu n}}$	n	$\frac{\|a_n - a^\dagger\|}{2^{-\nu n}}$	n	$\frac{\|a_n - a^\dagger\|}{2^{-\nu n}}$
5	3.1733	5	6.7243	5	8.9582
6	0.6096	6	6.3623	6	1.2236
7	0.4926	7	9.0426	7	0.9587
8	0.4619	8	5.6095	8	0.3386
9	0.4637	9	5.3524	9	0.2489
10	0.4188	10	2.1440	10	0.0832

5 Conclusions and Remarks

The considerations and principles of Section 2 could probably be applied not only to finite elements but also to boundary elements or to wavelets as basis functions ψ_j^n.

To emphasize once more the fundamental difference to the use of singular functions as basis functions, we point out that in the latter case the eigenvalues of the matrix $M^n = (\langle A_n^*\psi_i^n, A_n^*\psi_j^n\rangle_{1\le i,j\le k(n)})$ are the singular values $\sigma_1^n, \ldots \sigma_{k(n)}^n$ of A_n, so that the largest eigenvalue is about the same for all $n \in \mathbb{N}$ and bounded away from zero, while the smallest one goes to zero as $n \to \infty$ (if A_n is compact). In contrast to this, the eigenvalues of the matrix M^n with the ψ_j^n being finite elements are all about the same size and all go to zero as $n \to \infty$ with the same rate, namely $O((h_n)^{2p})$.

For an implemetation of method (21) one needs to have the functions $A_n^*\psi_j^n$, $j = 1, \ldots k(n)$, and the matrix M^n. Since the basis functions ψ_j^n are "simple" functions (piecewise polynomial with small support) these quantities can — not only in the model example of Section 3 but also for more complex and higher dimensional problems — usually be calculated explicitely, possibly using symbolic methods.

For more complicated geometries in higher dimensions it would be desirable to do the mesh refinement in each Newton step adaptively instead of using uniform grids with a priori chosen grid size; due to the instability, one would have to make sure that the refinement does not proceed too fast (cf. condition (28)).

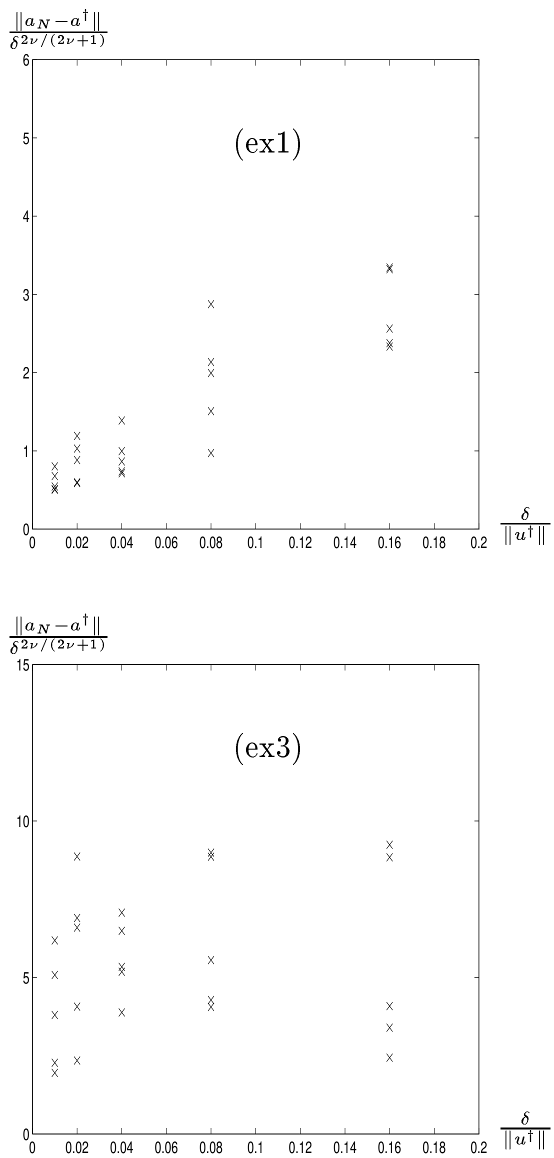

Figure 1: Results of a test series with five tests for each (relative) noise level $\delta/\|u^\dagger\| = 0.01, 0.02, 0.04, 0.08, 0.16$ and each example (ex1), (ex2), (ex3).

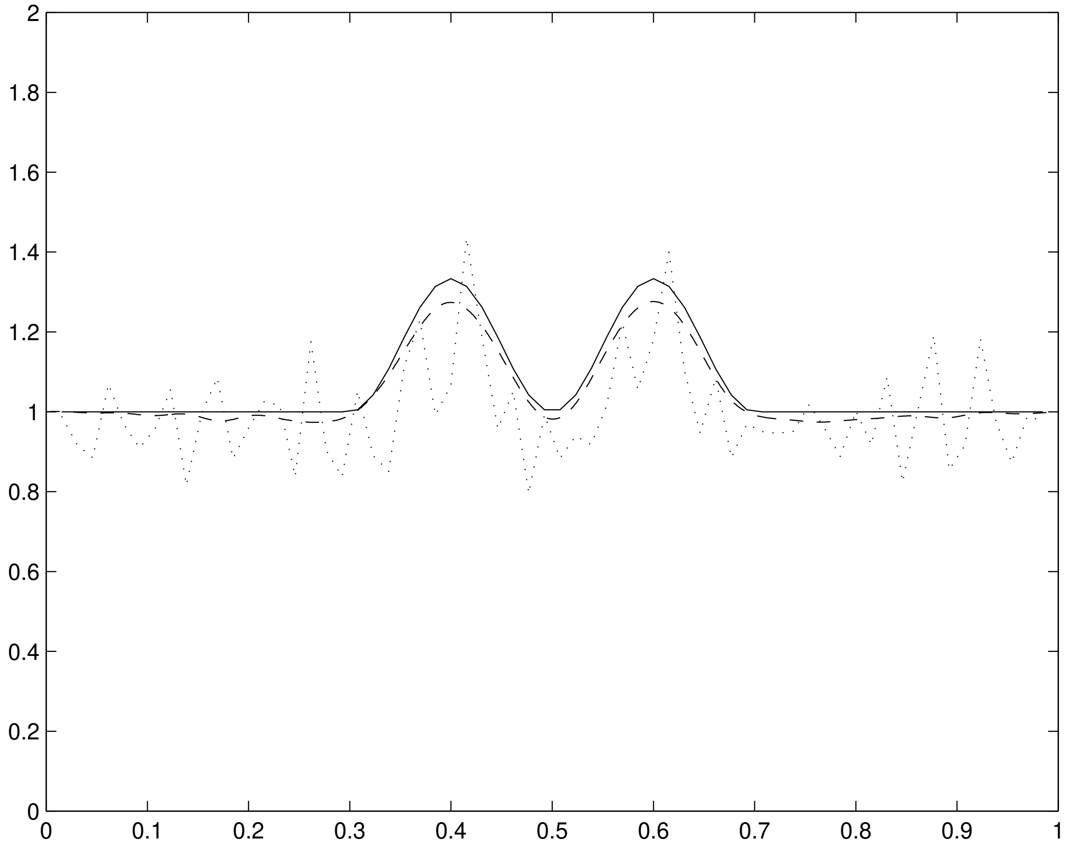

Figure 2: Approximations produced by Newton's method without regularization (dotted line: 3rd iterate) and with regularization by projection on finite elements of first order (dashed line: 10th iterate) from noisy data with noise level $\delta = 10^{-3}$, compared to the exact parameter a (solid line).

Appendix

Here we give the proofs of Section 3:

Proof.(Lemma 2)

Let $b \in (H^1(0,1))^*$ and let (b_n) be a sequence of $H_0^1(0,1)$-functions converging in $(H^1(0,1))^*$ to b. Then we define $F'(a)b$ to be the $L^2(0,1)$-limit as $n \to \infty$ of $v_n := F'(a)b_n$. We show that this limit exists and is independent of choice of the approximating sequence (b_n) and that, with some constant $C_{F'(a)}$ independent of b,

$$\|F'(a)b\|_{L^2(0,1)} \leq C_{F'(a)} \|b\|_{(H^1(0,1))^*} \tag{56}$$

holds: For arbitrary but fixed $m, n \in \mathbb{N}$ one has that $\tilde{v} := v_n - v_m$ solves

$$\begin{array}{ll} \tilde{v} \in H_0^1(0,1) & \wedge \\ \forall V \in H_0^1(0,1) : & \int_0^1 a \cdot \tilde{v}_x \cdot V_x \; dx = \int_0^1 (b_n - b_m) \cdot F(a)_x \cdot V_x \; dx \end{array}$$

Let now $\tilde{V}$ be the solution of

$$\begin{array}{ll} \tilde{V} \in H_0^1(0,1) & \wedge \\ \forall W \in H_0^1(0,1) : & \int_0^1 a \cdot \tilde{V}_x \cdot W_x \; dx = \int_0^1 \tilde{v} \cdot W \; dx \; . \end{array}$$

Then $\|\tilde{V}\|_{H^2} \leq C\|\tilde{v}\|_{L^2}$ with some $C > 0$ independent of $\tilde{v}$ and therefore

$$\begin{array}{rcl} \|\tilde{v}\|^2_{L^2(0,1)} & = & \int_0^1 a \cdot \tilde{V}_x \cdot \tilde{v}_x \\ & = & \int_0^1 (b_n - b_m) \cdot F(a)_x \cdot \tilde{V}_x \\ & \leq & 2\|\mathrm{id}\|_{H^1(0,1)\to L^\infty(0,1)} \|b_n - b_m\|_{(H^1(0,1))^*} \|F(a)\|_{H^2(0,1)} \underbrace{\|\tilde{V}\|_{H^2(0,1)}}_{\leq C\|\tilde{v}\|_{L^2(0,1)}} \; , \end{array} \tag{57}$$

which, with $C_{F'(a)} := 2C\|\mathrm{id}\|_{H^1(0,1)\to L^\infty(0,1)} \|F(a)\|_{H^2(0,1)}$ yields (56) with $F'(a)b$ replaced by $v_n - v_m$ and b by $b_n - b_m$. So (v_n) is a Cauchy sequence and therefore convergent; taking limits yields (56). □

Proof.(Lemma 3)

To show the second part of the Lemma we observe that for arbitrary but fixed $\tilde{w} \in L^2(0,1)$ and any $V \in H^1(0,1)$ one obtains, by using the symmetry of the operators $A(1)^{-1}$, $A(a)^{-1}$ with respect to the L^2-inner product as well as integration by parts together with the fact that $(A(a)^{-1}\tilde{w})(0) = (A(a)^{-1}\tilde{w})(1) = 0$:

$$\begin{array}{rcl} \int_0^1 F'(a)^*\tilde{w} \cdot V \; dx & = & -\int_0^1 A(1)^{-1}[F(a)_x \cdot (A(a)^{-1}\tilde{w})_x] \cdot V \; dx \\ & = & -\int_0^1 F(a)_x \cdot (A(a)^{-1}\tilde{w})_x \cdot A(1)^{-1}V \; dx \\ & = & \int_0^1 A(a)^{-1}\tilde{w} \cdot (F(a)_x \cdot A(1)^{-1}V)_x \; dx \\ & = & \int_0^1 \tilde{w} \cdot A(a)^{-1}[(F(a)_x \cdot A(1)^{-1}V)_x] \; dx \\ & \leq & \|\tilde{w}\|_{(H^4(0,1))^*} \underbrace{\|A(a)^{-1}[(F(a)_x \cdot A(1)^{-1}V)_x]\|_{H^4(0,1)}}_{\leq C_{F'(a)^*}\|V\|_{H^1(0,1)}} \end{array}$$

for some constant $C_{F'(a)^*} > 0$. Now, similarly to the proof of Lemma 2, a Cauchy sequence argument, as well as a limiting argument yield the assertion.

The proof of (52) is also done by using similar estimates and taking limits in the respective spaces. □

Proof.(Lemma 4)

Decomposing the $(H^{l+p})^*$ norm of $W := P_{\mathrm{span}\{\psi_1^n,\dots,\psi_{k(n)}^n\}^\perp} w$ into the respective portions on the elements $\Omega_i = [(i-1)h_n, ih_n]$, $i = 1, \frac{1}{h_n}$, we get, using the Riesz representation $\bar{W} \in H^{l+p}(0,1)$ of W, i.e., a $\bar{W}$ such that $W(v) = \langle \bar{W}, v\rangle_{H^{l+p}(0,1)}$, $\forall v \in H^{l+p}(0,1)$, that

$$\|W\|^2_{(H^{l+p}(0,1))^*} = \sum_{j=1}^{\frac{1}{h_n}} \|W\|^2_{(H^{l+p}(\Omega_j))^*} \tag{58}$$

Moreover we make use of the fact that for the transformation $\tilde{\phi}$ of some $\phi \in H^k(\Omega_i)$ on the reference element $K = (0,1)$ (not to be mischanged with the domain $(0,1)$ in (48)!) one has

$$(h_n)^{-m-\frac{1}{2}} \|\phi\|_{H_0^m(\Omega_i)} = \|\tilde{\phi}\|_{H_0^m(K)} \ , \tag{59}$$

for any $m \in \mathbb{N}$. Since we have to estimate an $(H^m)^*$ norm (and not an $(H_0^m)^*$ norm) the Bramble-Hilbert Lemma will be useful:

Lemma 6. *(Bramble-Hilbert)*
Let $D \subseteq \mathbb{R}^d$ be a domain, $m \in \mathbb{N}$, and let Q be a linear continuonus functional on D, with

$$\forall p \in P_{m-1} \ : \quad Q(p) = 0 \ , \tag{60}$$

where P_{m-1} denotes the polynomials of degree less or equal to $m-1$.
Then there exists a constant C_D depending only on the domain D, such that

$$\forall u \in H^m(D) \ : \quad |Q(u)| \le C_D \|Q\|_{(H_0^m(D))^*} \|u\|_{H_0^m(D)} \ ,$$

where $\|v\|_{H_0^m(D)} := \sqrt{\sum_{|\alpha|=m} \|D^\alpha v\|_{L^2(D)}}$.

*Proof.*see e.g. Lemma 4.8 and its proof in Section 4.5 in [14].

By inspection of the proof of Lemma 4.8 in [14] one can — at least in the one-dimensional case — easily show that $C_D = O(\max\{1, \sqrt{\mathrm{meas}(D)}\})$.

Setting $Q := W$ and using the Riesz representation $\bar{W} \in H^l$ of W, one easily sees that (60) holds with $m = l$. This, together with (59) yields

$$\begin{aligned}
\|W\|_{(H^{l+p}(\Omega_j))^*} &= \sup_{\phi \in H^{l+p}(\Omega_j)\, \phi \neq 0} \frac{W(\phi)}{\|\phi\|_{H^{l+p}(\Omega_j)}} \\
&\le \sup_{\phi \in H^{l+p}(\Omega_j)\, \phi \neq 0} \frac{W(\phi)}{\|\phi\|_{H_0^{l+p}(\Omega_j)}} \\
&= h^{l+p} \sup_{\tilde{\phi} \in H^{l+p}(K)\, \tilde{\phi} \neq 0} \frac{\tilde{W}(\tilde{\phi})}{\|\tilde{\phi}\|_{H_0^{l+p}(K)}} \\
&\le h^p \tilde{C} \sup_{\phi \in H^{l}(\Omega_j)\, \phi \neq 0} \frac{W(\phi)}{\|\phi\|_{H_0^{l}(\Omega_j)}} \\
&\le h^p \tilde{C} C_{\Omega_j} \|W\|_{(H_0^l(\Omega_j))^*} \ ,
\end{aligned}$$

where $\tilde{C} = \dfrac{\sup_{\tilde{\phi} \in H^{l+p}(K)\, \tilde{\phi} \neq 0} \frac{\tilde{W}(\tilde{\phi})}{\|\tilde{\phi}\|_{H_0^{l+p}(K)}}}{\sup_{\tilde{\phi} \in H^{l}(K)\, \tilde{\phi} \neq 0} \frac{\tilde{W}(\tilde{\phi})}{\|\tilde{\phi}\|_{H_0^{l}(K)}}}$ and C_{Ω_j} can be uniformly bounded by a constant C independent of h_n. Now by (58) the assertion follows. □

Proof.(Lemma 5)

With our special choice of A_0 we have, for any $w \in Y$ (setting the factor $\frac{g^1-g^0}{a^0}$ equal to one for simplicity and w.o.l.o.g.)

$$A_0 A_0^* w = -A(1)^{-1}[(A(1)^{-1}[(A^{-1}w)_x])_x] \ ,$$

i.e., $A_0 A_0^* w =: v$ is the solution of the fourth order boundary value problem

$$\begin{array}{rcll} v_{xxxx} & = & w & \text{in } (0,1) \\ v(0)=v(1) & = & 0 & \\ v_x(0) & = & v_x(1) & \\ v_{xx}(0) & = & v_{xx}(1) & \end{array} \tag{61}$$

Instead of the matrices $M^n := [\langle v_i^n, \psi_j^n\rangle]_{1\le i,j\le k(n)}$, $S^n := [\langle v_i^n, v_j^n\rangle]_{1\le i,j\le k(n)}$, where $v_j^n := A_0 A_0^* \psi_j^n$, we consider the transformed matrices $\hat{M}^n = TM^n$, $\hat{S}^n := TS^nT$, where T is the $k(n)\times k(n)$-matrix with the entries

$$T_{i,j} = \begin{cases} 6 & \text{if } j=i \\ -4 & \text{if } j = i\pm 1 \\ 1 & \text{if } j = i\pm 2 \\ 0 & \text{else} \end{cases} \quad i \in \{2,\dots,k(n)-1\},\ j\in\{1,\dots,k(n)\}$$

$$\begin{array}{ll} T_{1,1} = T_{k(n),k(n)} = \frac{56}{5} & T_{1,k(n)-3} = T_{k(n),4} = -\frac{3}{40} \\ T_{1,2} = T_{k(n),k(n)-1} = -\frac{36}{5} & T_{1,k(n)-2} = T_{k(n),3} = \frac{8}{15} \\ T_{1,3} = T_{k(n),k(n)-2} = \frac{32}{15} & T_{1,k(n)-1} = T_{k(n),2} = -\frac{9}{5} \\ T_{1,4} = T_{k(n),k(n)-3} = -\frac{7}{40} & T_{1,k(n)} = T_{k(n),1} = \frac{24}{5} \\ T_{1,j} = T_{k(n),k(n)+1-j} = 0 & j \in \{5,\dots,k(n)-4\} \end{array}$$

Note that $(h_n)^{-4}T$ is just the discretization of the boundary value problem (61) by finite differences. Indeed, using the fact that the ψ_j^n are piecewise linear and that $\psi_{j+l}^n(x) = \psi_j^n(x - lh_n)$ for any integer l, one can show that

$$T \begin{pmatrix} v_1^n \\ \vdots \\ v_{k(n)}^n \end{pmatrix} = (h_n)^4 \begin{pmatrix} \psi_1^n \\ \vdots \\ \psi_{k(n)}^n \end{pmatrix} \ .$$

Therefore

$$\hat{M}_{i,j}^n = (h_n)^4 \langle \psi_i^n, \psi_j^n\rangle \ , \quad \hat{S}_{i,j}^n = (h_n)^8 \langle \psi_i^n, \psi_j^n \rangle \ ,$$

so the transformed matrices $\hat{M}^n$, $\hat{S}^n$ are tridiagonal and their eigenvalues can be easily estimated, using Gershgorin's theorem. Summarizing we obtain

$$\begin{array}{rcl} \sup_{\alpha\in\mathbb{R}^{k(n)}} \frac{\alpha^T S^n \alpha}{\alpha^T (M^n)^2 \alpha} & = & \sup_{\alpha\in\mathbb{R}^{k(n)}} \frac{(T^{-1}\alpha)^T \hat{S}^n T^{-1}\alpha}{(T^{-1}\alpha)^T \hat{M}^n (\hat{M}^n)^T T^{-1}\alpha} \\ & \le & \frac{\lambda_{\max}(\hat{S}^n)}{\lambda_{\min}(\hat{M}^n(\hat{M}^n)^T)} \\ & \le & \frac{3/2(h_n)^8}{(1/2(h_n)^4)^2} = 6 \ . \end{array}$$

To estimate $\lambda_{\min}(M^n)$ from below, we also consider $\hat{M}^n$ and obtain

$$(\lambda_{\min}(M^n))^2 \ge \frac{\lambda_{\min}(\hat{M}^n(\hat{M}^n)^T)}{\lambda_{\max}(T^2)} \ge \frac{(1/2(h_n)^4)^2}{(335/12)^2} \ ,$$

where we have used the fact that $(\lambda_{\min}(M^n))^2 \ge \lambda_{\min}((M^n)^2)$ in the first inequality. Now an application of Lemma 1 yields the assertions. $\square$

Acknowledgement

The author wishes to thank Prof. Martin Hanke for his valuable comments on a first version of this paper.

References

[1] A.B. Bakushinskii, *The Problem of the convergence of the iteratively regularized Gauss–Newton method*, Comput.Math.Math.Phys. 32 (1992), 1353–1359.

[2] B. Blaschke(-Kaltenbacher), *Some Newton type methods for the solution of nonlinear ill-posed problems*, PhD thesis, Schriften der Johannes Kepler Universität Linz, Trauner, Linz, 1996.

[3] B. Blaschke(-Kaltenbacher), A. Neubauer, and O. Scherzer, *On convergence rates for the iteratively regularized Gauß–Newton method*, IMA J. Numer. Anal. 17 (1997), 421–436.

[4] J.H. Bramble, Z. Leyk, J.E. Pasciak, *The analysis of multigrid algorithms for pseudo-differential operators of order minus one*, preprint, Texas A& M University, 1995.

[5] G. Chavent, K. Kunisch *On weakly nonlinear ill–posed problems*, SIAM J. Appl. Math. 56 (1996), 542–572.

[6] P. Deuflhard, G. Heindl, Affine invariant convergence theorems for Newton's method and extensions to related methods, SIAM J. Numer. Anal. 16 (1979), 1–10.

[7] H.W. Engl, *On the convergence of regularization methods for ill-posed operator equations*, in: G. Hämmerlin, K.H. Hoffmann, eds., *Improperly Posed Problems and their Numerical Treatment*, Birkhäuser, Basel, 1983, 81–95.

[8] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996

[9] H.W. Engl, K. Kunisch, A. Neubauer, *Convergence rates for Tikhonov regularization of nonlinear ill–posed problems*, Inverse Problems 5 (1989), 523–540.

[10] H.W. Engl, A. Neubauer, *On projection methods for solving linear ill–posed problems*, in: A. Vogel, ed., Model Optimization in Exploration Geophysics, Vieweg, Braunschweig, 1987, 73–92.

[11] C.W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*, Pitman, Boston, London, Melbourne, 1984.

[12] C.W. Groetsch, *Inverse Problems in Mathematical Sciences*, Vieweg, Braunschweig, 1993.

[13] C.W. Groetsch, A. Neubauer, *Convergence of a general projection method for an operator equation of the first kind*, Houston J. Mathem. 14 (1988), 201–207.

[14] Ch. Grossmann, H.-G. Roos *Numerik partieller Differentialgleichungen*, Teubner, Stuttgart, 1994.

[15] M. Hanke, *A regularization Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems 13 (1997), 79–95.

[16] M. Hanke, *Regularizing properties of a truncated Newton–CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim. 18 (1997), 971-993.

[17] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math. 72 (1995), 21–37.

[18] M. Hanke, C.R. Vogel, *Two-level preconditioners for regularized inverse problems I: Theory*, Preprint Nr. 98/17, University of Karlsruhe, 1998.

[19] B. Kaltenbacher, *Some Newton type methods for the regularization of nonlinear ill–posed problems*, Inverse Problems 13 (1997), 729–753.

[20] B. Kaltenbacher, *A posteriori parameter choice strategies for some Newton type methods for the regularization of nonlinear ill–posed problems*, Numer. Math. 79 (1998), 501-528.

[21] J.T. King, *Multilevel algorithms for ill-posed problems*, Numer. Math. 61 (1992), 311–334.

[22] A. K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

[23] V.A. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, 1993.

[24] M.Z. Nashed and X. Chen, *Convergence of Newton-like methods for singular operator equations using outer inverses*, Numer. Math. 66 (1993), 235–257.

[25] F. Natterer, *Regularisierung schecht gestellter Probleme durch Projektionsverfahren*, Numer. Math. 28 (1977), 329–341.

[26] F. Natterer, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.

[27] A.N. Tikhonov, V.A. Arsenin, *Methods for Solving Ill-Posed Problems*, Nauka, Moskau, 1979.

[28] G.M. Vainikko, A.Y. Veterennikov, *Iteration Procedures in Ill–Posed Problems*, Nauka, Moscow, 1986. In Russian.